関西学院大学　先端社会研究所紀要　第１号
Annual Review of the Institute for Advenced Social Research vol.1

■論 文■

社会調査データとしての新聞記事の可能性
―― 読者投稿欄の計量テキスト分析試論 ――

中　野　康　人
（関西学院大学社会学部）

■要　　旨■　本稿の目的は、新聞記事データを社会調査データとして取り扱う準備作業を紹介し、その方法と問題点を整理することにある。メディア研究などを中心にして、新聞記事データの分析は長年の蓄積がある。しかし近年のデータベースの整備や、分析手法の発展により、電子媒体上の日本語新聞記事データを量的に分析することが容易になってきた。新聞記事からは、ある事象の頻度や分布、さらには意味や文脈といったことを分析できる。また、読者投稿欄などを分析対象とすれば、人々の意見や世論の断片を探ることが可能になるだろう。本稿では、朝日新聞の記事データベースから「声」欄の記事のみを抽出し、分析する過程を紹介する。「声」には、職業や年齢が明記されており、そうした属性と記事内容との関連を分析することが期待される。2006年の記事を具体例として、年齢や職業の分布、記事内容の概要を紹介する。

■キーワード■　新聞記事、計量テキスト分析、内容分析、世論調査

1　問題の所在

本稿の目的は、新聞記事データを社会調査データとして取り扱う準備作業を紹介し、その方法と問題点を整理することにある。

分析の詳細に入る前に、なぜ新聞記事データなのか、まずこの点を確認しておきたい。その理由の一つは、社会調査をとりまく状況にある。社会を研究対象とし、その実証手段として社会調査に携わる者であれば、社会調査をとりまく危機的状況を憂慮しない者は少ないだろう。その憂慮の一因は、社会調査、特に調査票調査の実施が困難になってきている、ということである。調査環境の問題としては、標本抽出の要となる各種リストの閲覧が困難になってきているということ、個人情報やプライバシーに対する意識の高まりもあり調査票の回収率が下がってきていることなどがあげられる（盛山2008，大谷2008 など）。また、統計学的な分析の方法論の視点からの問題点の指摘もある（Winship and Sobel 2004）。一般的な社会調査から得られるデータは観察データもしくは非実験データであり、メカニズムや因果のプロセスを推定するためのデータとしては制約が多い、という主張である。こうした問題意識を出発点とすれば、標本に基づく調査票調査という既存の社会調査の枠組みにとらわれず、様々なタイプの調査方法やデータ収集方法に挑戦することが、社会調査を行う者に求められているといえるだろう。

本稿は、そうした試みの一つとして、新聞記事データを社会調査データとして扱うことを意図した作業の第一段階を紹介するものである。

2 新聞記事の分析

新聞記事を分析の対象とすること自体は、社会学においてさして新しいものではない。むしろ、分析の情報源として古くから重宝されてきたともいえる。なぜなら、新聞には定期的に社会の出来事が記述され、そしてそれが蓄積されていっており、現在の社会状況を映す鏡であると同時に、過去の状況を教えてくれる貴重な資料でもあるからだ[1)]。

特に、内容分析（content analysis）の方法論が確立した1950年代以降、新聞記事の分析は社会分析の一つの常套手段であった。メディア研究、コミュニケーション研究といった文脈では、新聞そのものが研究の対象であり、その研究の蓄積は枚挙に暇が無い。Neuendorf（2002, pp.28-29）は、主要な社会科学雑誌に掲載された内容分析の論文の数をまとめている。たとえば、Sociological Abstracs では、1960年代には内容分析をキーワードに含む論文は毎年一桁であったが、漸次論文数は増加し、1980年代後半以降は常に三桁の論文数を維持している。日本でも、メディア研究の蓄積は豊富である。それ以外の流れで新聞記事を社会調査データとして使用した例は、見田（1965）が有名であろう。見田は、読売新聞に掲載された身の上相談を用いて「不幸の類型」を試みている。

新聞記事を社会調査データとして使うということは、調査票調査に基づいた社会調査データと何が異なるのか。一般の調査票調査では、自らの知りたいことを操作化して質問文に変換し、その質問文に対する回答者の反応をデータとする。計量テキスト分析では、自らの知りたいことをテキストから抜きだして集計することでデータを得る。研究関心に適合的なテキストを準備し、そこからデータを得る作業は、調査票調査で調査を実施してデータを得る作業と同等のものである。データを得るプロセスについては、標本抽出やデータの信頼性・妥当性・再現性・代表性の問題など、新聞記事固有の問題もあるが、調査票調査でも同じような問題が生じるといっても過言ではない。

方法論的には、内容分析、就中、新聞記事分析は、これまでどちらかといえば「質的な研究方法」に分類されてきた[2)]。あるテキスト事例の解釈という作業が中心的であったがゆえの分類であったのだろう。もちろん、記事の出現回数や範囲などを計数する「計量的な研究方法」も古くから存在した（Woodward 1934）しかし、記事が電子媒体で蓄積され、さらには文書を処理する情報技術が発展したことと相まって、近年は計量的な分析が台頭してきている。樋口（2004a, 2004b）は、自ら分析用ツールを開発しつつ、新聞記事の分析に取り組んでいる。比較的新しい内容分析のテキストである Neuendorf（2002）では、内容分析のまとめ方の手法として、対応分析やパス解析などの分析手法が紹介されいている。Krippendorff も第二版では、計量的分析手法に大きな改訂が見られる（Krippendorff 1900, 2002）。

1）もちろん、記事になったことが社会のすべてだとか、新聞記事を読めば社会のすべてがわかるとか、新聞記事が客観的・中立的に書かれたものだ、などという愚かな主張をするつもりは毛頭無い。

2）例えば、質的分析の入門書である Grbich（2007）は、Content analysis of texts に一章を割いている。

イラン、前線に新兵力10万人２、３月に新作戦説も

【テヘラン五日＝吉田（秀）特派員】イラン・イラク戦争はペルシャ湾でのタンカー攻撃とともに、国境地帯でもイラク側の攻勢が強まり、軍事筋によるとこれに対抗するためイラン軍は五日までに新たな兵力約十万人を前線に送った。＜後略＞（朝日新聞1985年１月６日朝刊）

図１　事実報道の記事

真の軍縮を目指して（社説）

「いま一発の核爆発が欧州で起これば、軍の通信指揮系統がまっさきに破壊され、結局は全面核戦争に進まざるをえなくなる」「核戦争のあと黒い雨が降り、気温が低下し、地球は凍結する」「第三次大戦は、間違いなく人類最後の戦争となるはずだ」＜後略＞（朝日新聞1985年１月１日朝刊）

図２　社説記事

新聞記事を、メディア研究のデータとしてではなく、社会で発生している事実や人々の考えを分析するデータ、つまりは社会調査データとして使用するには、一定の制約が伴う。それは、新聞記事は社会のどのような部分を表象しているのか、ということである。新聞記事は、決して社会のすべての出来事を、中立的に表現・記録したデータとは言えないだろう。そこにあるのは、新聞社の報道傾向、記事を執筆した記者の主観、新聞読者側のニーズなど、様々なフィルターを通った上で表現された社会的事実である。また、報道が世論を形成するのか、それとも世論に基づいた報道がなされるのか、という疑問もある[3)]。しかし、詳細な議論はメディア論の分野でなされており、本稿の範囲をこえるものであるので、これ以上は踏み込まないでおく。

また、新聞記事の内容にはいくつかの種類がありえる。一つには、様々な事件や事象を伝える記事である（図１）。この種の記事には、どのような出来事が発生しているのかを、比較的客観的に伝える内容が含まれている。ある社会現象が、いつどこでどのように発生しているのかを知るデータとしては、この種の記事が重宝する。もちろん、厳密には、記者の主観や執筆時の社会通念を反映しているものである。それゆえ、こうした記事の分析から、社会に存在するステレオタイプの研究などもおこなえる。

こうした客観的事実報道とは別に、主観的な評価や意見を前面に押し出した記事もある（図２）。社説や投稿欄などがそれにあたる。そうした記事では、人々のもしくは社会の意見や価値観を表す言葉が出てくるので、世論や価値観の調査データとして扱いうる。もちろん、そこで表現されているのは、執筆者個人の主観だけでなく、編集者や新聞社のフィルターもかかっていることはいうまでもない。

そうした制約条件を留保しながら、目的に応じて、利用する記事を絞り込む必要があるだろう。

3）報道内容と世論調査結果を比較した研究もある（Hertog and Fan 1995）。しかしこれは、一概にいえるものではなく、トピックに応じてどちらの場合もありうるだろう。

3 テキストデータの取扱い

次に、日本語テキストを計量的に扱うことの難点を指摘しておく。

新聞記事データの分析には、かつては膨大な手作業が必要であった。紙媒体に印字された文字を拾い上げ、数を数え上げたり、電子媒体に入力したりという作業である。それゆえに、新聞記事データ自体は膨大な量があっても、分析に使用されるのはそのうちのごく一部の標本のみ、ということがよくあった。先述の見田（1965）が分析対象としたのは、一年分304件の身の上相談記事である。一方、太郎丸（1999）も、1934年、1964年、1994年それぞれから約100件ずつ、合計343件の記事を系統抽出で抜き出して分析している。

ここ10年ほどは、新聞記事データのデータベースが急速に整備され、電子媒体上で記事を入手することが容易になってきた。このことにより、記事の検索や分析の労力が大きく軽減され、記事データの利用が身近になってきた。例えば、近年の朝日新聞の記事は一年間に15万件ほど（一日平均400件超）である。人力でこれらの記事を抜き出して計数するのは非現実的であるが、電子媒体上であればたやすいことである。欧米で出版されている内容分析のテキストには、データベースやコンピュータを利用した分析の仕方が詳述されている（例えば、Neuendorf 2002）。しかし、日本に関しては事態は同様ではない。それは、日本語という言語の特質に由来する問題があるからである。

例えば、「戦争」という言葉が分析の対象である、と明確に決まっているのであれば話は簡単である。記事のテキストデータから、その言葉を検索して抽出することは容易である。しかしながら、ターゲットとなる言葉が明確にわかっていないときはどうだろうか。たとえば、戦争という言葉がどのような文脈で使われているのか、を知りたいとする。このとき、戦争と同時に出てくる言葉を抽出することによって、戦争が用いられる意味や文脈を分析する訳だが、ある日本語の文章から、そこに書かれている任意の単語を抽出する作業は、かなり困難な作業である。例えば英語であれば、単語と単語の間には必ず空白が存在するので、文章から単語を抽出するのは機械的に容易である。しかし、日本語の文章には単語間に決まった区切りがあるわけではないので、単語を抽出するには文章内に存在する言葉の意味を理解する必要がある。日本語を理解する人間であれば、その作業は可能である。したがって、人力でテキスト分析を行うのであれば問題はない。しかし、コンピュータで自動的に分析を行うとなると、そうはいかない。日本語という言語の特質が、機械的にテキストを分析する際の障壁になっていたのである。この問題を解決してくれたのが、近年の言語学・情報科学における自然言語処理の発展である。それは、形態素解析と呼ばれる技術で、日本語の文章から意味のある単語を切り分けて抽出してくれるというものである[4)]。

次のような文があるとする。

１．イラン・イラク戦争が勃発した。
２．新たな戦争が勃発する脅威がある。

４）形態素解析をしてくれるソフトウエアは、市販されているものもあれば、フリーで公開されているものもある。詳細は、藤井ら（2005）、石田（2008）などを参照。

この文は、次のような単語に分かち書きできる。

1．イラン／・／イラク／戦争／が／勃発した。
2．新たな／戦争／が／勃発する／脅威／が／ある。

ここから次のようなデータができる。

	戦争	勃発	イラン	イラク	新たな	脅威
文1	1	1	1	1	0	0
文2	1	1	0	0	1	1

このように、文書と単語の情報が数値化できれば、あとは調査票調査のデータと同じく、計量的なデータ分析の土俵にのることになる。日本語テキストを分析する技術は、近年になって整備され、一般に利用できるようになってきたと言えるだろう。社会調査としてだけではなく、マーケティング分野や情報科学分野で、テキストマイニングという呼称で、技術の蓄積がなされている。例えば、大塚ら（2007）、那須川（2006）、大隅（2002）、石田（2008）、藤井ら（2005）などを参照のこと。

では、新聞記事から上記のようなデータができたときに、そこからどのような分析ができるのであろうか。一般的な事実報道の記事においては、次のようなことが可能であろう。一つは、記事中に出現する言葉を計数することによって、ある社会的な出来事や概念がどのくらい発生・報道されているのかを知る、ということである。新聞記事が、時系列的に蓄積されていることを考慮すれば、その頻度の時間的変化をみることも可能である。表1は、朝日新聞について、「戦争」というキーワードを含む記事数を1985年と2002年のそれぞれにおいて月ごとに集計したものである(Nakano 2006)。これを見れば、戦争に関する報道は年間を通して一定数存在するが、特に八月においてその頻度が増すことがわかる。

新聞記事から分析できるもう一つのことは、記事中に出現する言葉と言葉の共起関係を集計する

表1　「戦争」を含む記事数の月ごとの変化（朝日新聞）

月	1985	2002
一月	42	263
二月	38	196
三月	38	246
四月	34	213
五月	52	294
六月	48	283
七月	36	277
八月	79	365
九月	50	234
十月	55	225
十一月	45	247
十二月	43	262
合計	560	3105

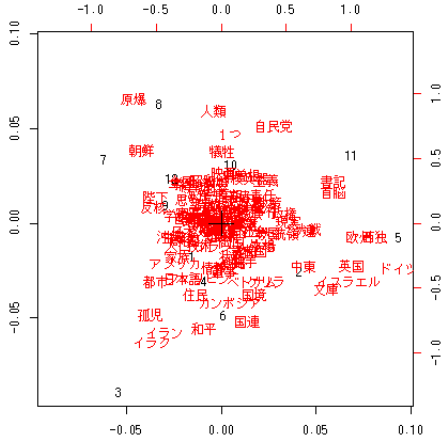

図３　戦争記事の文脈（1985年朝日新聞）

私の宝お守りにする母の手製足袋

〇〇〇〇　幼稚園長千葉県八千代市62

戦争が激しくなって、東京の空襲を避けるため郷里の富山に戻ったのは、1944年の冬でした。寒さをしのぐため、母は自分のもんぺの生地を使って私に足袋を作ってくれました。今でも、お守り代わりにその足袋をリュックに入れて持ち歩いています。裏側には、ほころびを何度も継ぎ足した跡があります。空襲警報を聞くたびに防空壕（ごう）に駆け込んだことがよみがえるようです。定年後に、世界８カ国で開催されるウオーキング大会に出て、国際マスターウオーカーを目指す計画を立てました。飛行機に乗らないと、日本以外の７カ国には行けません。ところが、私は飛行機が大の苦手。そこで、足袋を持参して天国にいる母に守ってもらい、何とか計画を達成できました。私は小学校の校長をしていた時から、いつもこの足袋を教え子たちに見せてきました。単なる形見の品ではありません。二度と戦争を起こしてはいけないと、子どもたちに語りかけるための、貴重な証拠の品なのです。（朝日新聞2006年１月１日朝刊：氏名部分のみ改編）

図４　朝日新聞の「声」

ことによって、ある出来事や事象が、どのような文脈で語られているのかを明確にするということである。これは、その言葉もしくは事象がその社会で持ちうる意味を明かにする作業と言ってもよい。計数的な分析と同じく、事例列的な視点を導入することによって、ある事象の社会的意味が時代によってどのように変化するかを明らかにできる。図３は、1985年の朝日新聞で戦争をキーワードに含む記事中に含まれる頻出名詞を月ごとに集計したデータから対応分析を行ったものである。この図を見れば、例えば、八月の戦争記事は原爆の文脈で語られていることが明確である。

４ 朝日新聞「声」の分析

ここからは、事実報道の一般記事ではなく、読者投稿欄について考える。読者投稿欄についても、先述の一般記事と同じく、言葉の計数や文脈の分析が可能である。しかし、読者投稿欄は一般記事にはない特徴がある。それは、投稿者の属性が明記されている、という点である。図４は、朝日新

```
\ＩＤ\０６５８１６４０
\ＡＦ\１９８８０１０１
\Ａ０\Ｎ
\Ａ１\Ｍ
\Ａ２\５面
\Ａ３\Ｔ
\Ａ４\４６８
\Ｔ１\ぜひ実らせて私の“ 設計図”（声・今年私の夢は）
\Ｔ２\　和光市　〇〇　〇〇（主婦　５７歳）
\Ｔ２\　高校生の息子がテレビを見ながら「ホラホラ小学生が上手に粘土をこねてるよ。お母さんが作っているようなのは、小学生でも作れるということ」と、明らかに私の趣味の陶芸作品を揶揄（やゆ）しての言葉。父親はそばでフッフッと笑っていた。カチン。今に見ておれ、ウーンとうならせてみせるからね。
\Ｔ２\　日だまりのこたつの上で、私は夢の設計図を書き始めた。今年こそはと闘志をわかせながら、胸は次第に熱をおび、ふつふつと煮えたぎってきた。
\Ｔ２\　夢の設計図とは、美術品集中展示室のこと。手持ちの陶芸の先生方の作品。好きな絵や版画、そしてちょっぴり自分の会心の作を並べられる棚、時には外に開放して……。でも夢の設計はすぐ境界にぶつかる。そうだ天へのびよ。狭い土地ゆえ、初めから３階に建てれば家族の核分裂もしないですむ。
\Ｔ２\　陶磁の世界は深く美しい。そこにひたれば怒りも悲しみも希望に変化しストレスも解消。背後より夫の声あり。「好き勝手してストレスもないだろう」だって。はいだんな様。感謝してます。だから夢の設計図、実らせてください。
```

図５　「声」の記事データ（氏名部分のみ改編）

聞の読者投稿欄である「声」の記事である。「声」には、投稿者の氏名、職業、住所、年齢が明記されている。この情報から、人々の属性と、そこから発せられる言葉の関係を分析することが可能になる。以下では、そうした分析をするための準備作業として、朝日新聞の「声」のデータを整理した過程を紹介する。

4.1 使用するデータ

今回使用するデータは、朝日新聞社が発行し日外アソシエーツが販売している「朝日新聞記事データ集　学術・研究用」である。このデータは、内容分析にかけられることを著作権者が認めた形で販売されているもので、朝日新聞本社版の記事テキストを基本的にすべて収録したものである。したがって、事実報道の一般記事も含めて分析の対象となりうる（中野2009）。しかしここでは、その中から「声」の記事のみを取り出して分析をする作業を扱う。

図５は、当該データ中に含まれる「声」の一例である。データは、タグによって区切られ、記事本文だけでなく、いくつかの記事情報も付随している（表２）。

「声」の記事は、「＼Ａ２＼」が「５面」もしくは「オピニオン」となっている[5)]。ただし、そ

５）データの格納の仕方や面名の表記が年によって微妙に変化している。場合によっては、年の途中で変化していることもある。分析者にとっては、非常に扱いづらいデータである。

表２　朝日新聞記事データのタグ

タグ	意味
＼ＩＤ＼	記事 ID
＼ＡＦ＼	年月日
＼Ａ０＼	紙誌名
＼Ａ１＼	朝夕刊
＼Ａ２＼	面名
＼Ａ３＼	発行者
＼Ａ４＼	文字数
＼Ａ５＼	分類
＼Ｔ１＼	記事タイトル
＼Ｔ２＼	記事本文

“title”, “year”, “month”, “day”, “occupation”, “address”, “age”, “article”
“ぜひ実らせて私の “設計図””, “1988”, “01”, “01”, “主婦”, “和光市”, “57”, “高校生の息子がテレビを見ながら「ホラホラ小学生が上手に粘土をこねてるよ。お母さんが作っているようなのは、小学生でも作れるということ」と、明らかに私の趣味の陶芸作品を揶揄（やゆ）しての言葉。父親はそばでフッフッと笑っていた。カチン。今に見ておれ、ウーンとうならせてみせるからね。日だまりのこたつの上で、私は夢の設計図を書き始めた。今年こそはと闘志をわかせながら、胸は次第に熱をおび、ふつふつと煮えたぎってきた。夢の設計図とは、美術品集中展示室のこと。手持ちの陶芸の先生方の作品。好きな絵や版画、そしてちょっぴり自分の会心の作を並べられる棚、時には外に開放して……。でも夢の設計はすぐ境界にぶつかる。そうだ天へのびよ。狭い土地ゆえ、初めから３階に建てれば家族の核分裂もしないですむ。陶磁の世界は深く美しい。そこにひたれば怒りも悲しみも希望に変化しストレスも解消。背後より夫の声あり。「好き勝手してストレスもないだろう」だって。はいだんな様。感謝してます。だから夢の設計図、実らせてください。”

図６　「声」の csv データ

の面には「声」以外の記事も含まれるので、さらに「＼Ｔ１＼」で「声」を含む記事を特定する必要がある[6)]。記事が特定できれば、そこから「年月日」「住所」「職業」「年齢」「タイトル」「記事本文」を抽出して、データ化できる[7)]。図６は、図５の記事データを csv 化したものである。中野は、すべての記事データから自動的に「声」の記事のみを抽出して csv 化するスクリプトを perl で作成し、データを抽出した[8)]。これで、年月日、住所、職業、年齢、なんらかの意見表明をしている投稿記事を含んだ社会調査データができることになる。

６）ただし、これは必要条件だが十分条件ではない。「声」以外の記事でもタイトルに「声」という文字が含まれることがある。例えば、「首脳の共同声明」など。厳密には、「(声)」がタイトルの行頭にあるとか、「(声)」が行末にあるなどといった条件でデータを抽出する必要がある。

７）「住所」「職業」「年齢」の書式は、年によってまちまちであり、データ抽出には注意が必要である。また、「名前」の抽出も可能だが、個人の特定や名前による分析を目的とはしないので、名前はデータの中に含めない。

８）スクリプトの詳細は、中野に問い合わせられたい。

5 2006 年の「声」

ここからは、実際のデータを簡単に分析して、「声」がどのような投稿者によってなりたち、どのようなテーマが主に取り上げらているかを確認する。2006年の朝日新聞記事データに含まれる総記事数は、154385件である。そのうち、前節の手順を踏んで抽出される「声」の記事は2526件で、全記事の約1.6%を閉める。表３は、日ごとの「声」の記事件数一覧である。新聞休刊日などをのぞいて、基本的に毎日６から９件の掲載がある。

表３　「声」の日ごとの件数（2006年）

月	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	9	0	9	9	9	8	7	8	7	7	8	7	7	7	8	8	7	7	7	7	7	8	8	7	0	7	7	7	8	6	7
2	8	7	7	7	8	8	7	8	7	7	7	8	0	7	8	7	7	7	8	8	7	0	7	7	7	8	8	7	0	0	0
3	5	6	7	7	7	8	7	8	7	7	7	8	8	7	8	7	7	7	6	8	7	0	7	7	7	8	8	7	6	7	7
4	7	8	8	7	8	7	7	6	8	0	7	8	7	7	7	8	0	7	8	7	7	7	8	8	7	8	7	7	6	8	0
5	6	7	7	7	7	0	8	8	7	8	7	7	7	8	0	7	8	7	7	7	8	8	7	8	7	7	7	8	8	7	6
6	7	7	7	8	7	7	8	7	7	7	8	8	7	8	7	7	7	8	0	7	8	7	7	7	8	4	7	6	7	7	0
7	6	7	8	7	8	7	7	7	8	0	7	8	7	7	7	8	0	7	8	7	7	7	8	8	7	8	7	7	7	8	6
8	7	8	7	7	7	8	7	7	8	7	7	7	8	0	7	8	7	7	7	8	0	7	8	7	7	7	8	8	7	6	7
9	7	7	8	8	7	8	7	7	7	8	0	7	8	7	7	7	8	8	7	8	7	7	7	8	8	7	6	7	7	7	0
10	8	8	7	8	7	7	7	8	8	0	8	7	7	7	8	8	7	8	7	7	7	8	8	7	8	7	7	7	8	6	7
11	8	7	7	7	8	8	7	7	7	7	7	8	0	7	7	7	6	7	8	8	7	8	7	7	7	8	8	7	6	6	0
12	7	7	8	8	7	8	7	7	7	8	0	6	8	7	7	7	8	8	7	8	7	7	7	8	8	7	6	7	4	8	6

図７は、投稿者の年齢分布である。投稿者の最年少は７才、最高齢は94才である。平均値は約53才、中央値は56才である。年齢分布には三つのピークがある。一つは60から70代の山、もう一つは40代の山、そして最後に10代の山である。逆にいえば、20代と50代に分布の落ち込みがある。

次に職業の分布を確認してみる（表４）。2006年の「声」投稿者には、335の異なる職業が確認される。そのうち、最も多いのは「無職」である。658件で全体の約26%を占める。次いで、「主婦」の530件（21%)、「会社員」の188件（７%）が続く。その後は、「高校生」（111件４%）「大学生」（94件４%）「中学生」（64件３%）である。図７とあわせて考えれば、投稿者の多数を占め

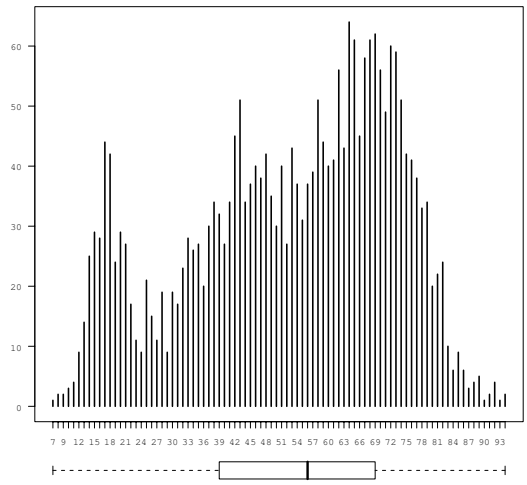

図７　「声」投稿者の年齢分布（2006年）

表４　「声」投稿者の職業（2006年）

職　業	件数	職　業	件数	職　業	件数	職　業	件数
無職	658	英会話非常勤講師	2	記録映画監督	1	生け花教室主宰	1
主婦	530	家庭教師	2	貴石加工業	1	生活サービス業	1
会社員	188	介護支援専門員	2	技術コンサルタント	1	精神保健福祉士	1
高校生	111	絵本作家	2	技術翻訳業	1	製造会社員	1
大学生	94	開業医	2	喫茶店経営	1	製品開発業	1
中学生	64	学習塾経営	2	居宅介護支援事業所役員	1	設計士	1
高校教員	49	看護学生	2	教育カウンセラー	1	専門学校非常勤講師	1
パート	35	救急科専門医	2	金物店経営	1	大学・福祉団体勤務	1
農業	32	教育相談員	2	靴店経営	1	大学院科目履修生	1
地方公務員	26	教員	2	経営コンサルタント	1	大学院教員	1
医師	22	広報コンサルタント	2	警備員	1	大学院聴講生	1
公務員	17	高校教師	2	建具業	1	大学院非常勤講師	1
小学校教員	17	高校教頭	2	建築家	1	大学受験生	1
会社役員	16	高校教諭	2	建築業	1	大学助教授	1
小学生	16	産婦人科医	2	建築整備士	1	短歌集団主宰	1
アルバイト	15	司法書士	2	建築設計	1	短大講師	1
大学非常勤講師	13	市議会議員	2	検針委託員	1	地域防災研究所長	1
団体役員	13	市職員	2	研究職	1	中学校教諭	1
大学院生	12	歯科医師	2	個人投資家	1	中学校司書	1
大学教員	12	歯科技工士	2	古物商	1	中学高校教員	1
看護師	11	社会教育指導員	2	語学講師	1	中学指導助手	1
団体職員	11	社会福祉士	2	公認会計士	1	朝鮮学校生	1
中学校教員	11	出版業	2	公民館長	1	町議会議員	1
塾講師	10	助産師	2	公立中教員	1	町職員	1
派遣社員	9	小学校非常勤講師	2	工業デザイナー	1	町臨時職員	1
弁護士	9	専門学校講師	2	校正業	1	調教師	1
翻訳業	8	貸しビル業	2	航空会社勤務	1	調理師	1
専門学校生	7	大学講師	2	高校事務職員	1	鳥取県原水協理事長	1
大学職員	7	非常勤講師	2	高校職員	1	通訳業	1
短大教員	7	病院職員	2	高専教員	1	庭園管理士	1
文筆業	7	福祉施設職員	2	高専生	1	添削指導員	1
予備校生	7	編集者	2	札幌市職員	1	電器店経営	1
ピアノ教師	6	弁理士	2	札幌市非常勤職員	1	島おこし団体役員	1
フリーター	6	保護司	2	仕立屋	1	東京都高校野球連盟審判委員長	1
高校講師	6	薬剤師	2	司法通訳	1	灯油販売業	1
大学教授	6	予備校講師	2	市議	1	独立行政法人研究職	1
留学生	6	養護学校教諭	2	市非常勤職員	1	内科医	1
会社経営	5	養護教諭	2	市民団体主宰	1	日本棋院普及指導員	1
学校事務職員	5	理学療法士	2	市民団体役員	1	日本語学校経営	1
国家公務員	5	鍼灸師（しんきゅうし）	2	市役所嘱託	1	日本語学校生	1
主夫	5	ＮＧＯ職員	1	指圧師	1	日本語講師	1
中学教員	5	ＮＰＯ職員	1	施設職員	1	日本語非常勤教師	1
保育士	5	たばこ小売り業	1	私塾教師	1	日本人学校教諭	1
１級建築士	4	まち美化ボランティア	1	紙工芸家	1	日本料理店主	1
アパート経営	4	エッセイスト	1	詩人	1	年金受給団体職員	1
フリー編集者	4	カトリック司祭	1	歯科衛生士	1	派遣研究員	1
ライター	4	カフェ経営	1	児童館職員	1	俳優	1
飲食業	4	カメラマン	1	児童文学作家	1	俳優訓練生	1
介護福祉士	4	ギャラリーオーナー	1	児童文学者	1	美術スクール講師	1
会社顧問	4	ケアマネジャー	1	写真家	1	百貨店レジ係	1
契約社員	4	コンビニ経営	1	社会福祉法人理事	1	病院嘱託	1
行政書士	4	システムエンジニア	1	社会保険労務士	1	病院役員	1
作家	4	スポーツ指導員	1	受験生	1	品質管理コンサルタント	1
専門学校教員	4	チェルノブイリ子ども基金理事	1	獣医師	1	不動産コンサルタント	1
著述業	4	ツアーコーディネーター	1	塾教師	1	不動産会社社長	1
日本語教師	4	デパート販売員	1	塾経営	1	不動産業	1
牧師	4	ドイツ語翻訳家	1	塾主宰	1	婦人服店経営	1
タクシー運転手	3	パート看護師	1	塾非常勤講師	1	福祉ワーカー	1
ペンション経営	3	ビオラ奏者	1	書籍販売業	1	文化団体役員	1
飲食店経営	3	ファイナンシャルアドバイザー	1	書店従業員	1	編集補助	1
園芸業	3	フリー学芸員	1	書道講師	1	保育カウンセラー	1
家事手伝い	3	ホームヘルパー	1	小学校教頭	1	保育パート	1
画家	3	ホームページ制作業	1	小学校校長	1	保育園長	1
会社社長	3	マンション管理士	1	小学校講師	1	保健師	1
学習塾講師	3	医学生	1	小学校事務職員	1	法科大学院生	1
技術士	3	医療事務員	1	小学校嘱託教員	1	法人理事	1
勤務医	3	印章彫刻士	1	小学校臨時職員	1	訪問介護員	1
高校非常勤講師	3	運送業	1	小規模作業所所長	1	無線技術士	1
漆器業	3	映画監督	1	小児科医	1	郵政公社職員	1
小児科医師	3	英会話教室主宰	1	小説家	1	郵便局員	1
整体師	3	園芸店経営	1	証券投資アドバイザー	1	用務員	1
税理士	3	横浜地裁判事	1	嘱託教員	1	養護学校教員	1
僧侶	3	音楽教師	1	職安職員	1	酪農会社員	1
短大生	3	音楽史研究家	1	職安相談員	1	理容師	1
中高教員	3	介護ヘルパー	1	職業	1	旅行ジャーナリスト	1
幼稚園長	3	介護士	1	職業訓練生	1	料理店経営	1
鍼灸師	3	介護施設職員	1	食品販売業	1	労組委員長	1
イラストレーター	2	介護職員	1	食料品店経営	1	労働組合役員	1
カウンセラー	2	介護認定調査員	1	新聞販売所員	1	労働団体役員	1
ジャーナリスト	2	会社嘱託	1	森林インストラクター	1	鍼灸（しんきゅう）師	1
タクシー乗務員	2	海外ボランティア	1	森林部門技術士	1		
ノンフィクション作家	2	外科医	1	神学生	1		
フリーライター	2	学校司書	1	診療所医師	1		
運転手	2	楽器商	1	診療所長	1		
英会話講師	2	楽器輸入卸業	1	図書館員	1		

るおおよその属性が想像できる。その他の職業については、複数登場するものもあるが、335の職業のうち210の職業は一回のみの登場である。ただし、この分類は記事に登場する職業の文字列をそのまま利用したもであるが、実質的には同じ「職業」と分類されるものが多数含まれる。

例えば、教員関係の職業名には次のようなものがある。

> 教員、非常勤講師、嘱託教員、小学校教員、小学校非常勤講師、小学校教頭、小学校校長、小学校講師、小学校嘱託教員、中学校教員、中学教員、中学校教諭、中高教員、中学高校教員、高校非常勤講師、高校教師、高校講師、高校教頭、高校教諭、養護学校教員、養護学校教諭、養護教諭、専門学校教員、専門学校講師、専門学校非常勤講師、語学講師、短大教員、短大講師、大学非常勤講師、大学教員、大学講師、大学助教授、大学教授、大学院教員、大学院非常勤講師

これら教育関係の職業を投稿者とする記事は179件あり、全体の約７％を占める。学校の種類による区別のみならず、その中の細かな職位や就業形態にまで言及しているものが少なからずある。また、臨時雇用の職業についても、各種「非常勤」の他に

> パート、アルバイト、フリーター、契約社員

といったものがある。こうした職業の提示の仕方も、記事内容とあわせて、ひとつの社会現象として分析の対象となりうるだろう。

次に、「声」の中身について簡単に触れておきたい。表５は、「声」のタイトルを形態素解析し名詞のみを取り出して、頻度を集計したものである。「世代」は時間を軸に前後を見据える内容を、「戦争」や「戦後」は戦争に関わる内容を、「法、首相、憲法」などは政治的な内容を、「子」や「教育」や「いじめ」は教育に関わる内容を、それぞれ表しているものと類推される。

表５　「声」タイトルにおける頻出単語（2006年）

単語	頻度	単語	頻度
世代	159	いじめ	34
私	102	言葉	34
年	90	首相	34
日	76	障害	31
戦争	73	憲法	30
人	68	新聞	30
者	67	国	29
心	52	父	29
夏	50	責任	28
法	48	米	28
教育	44	母	28
子	44	子供	27
戦後	41	氏	27
日本	40	化	25
必要	37	前	25

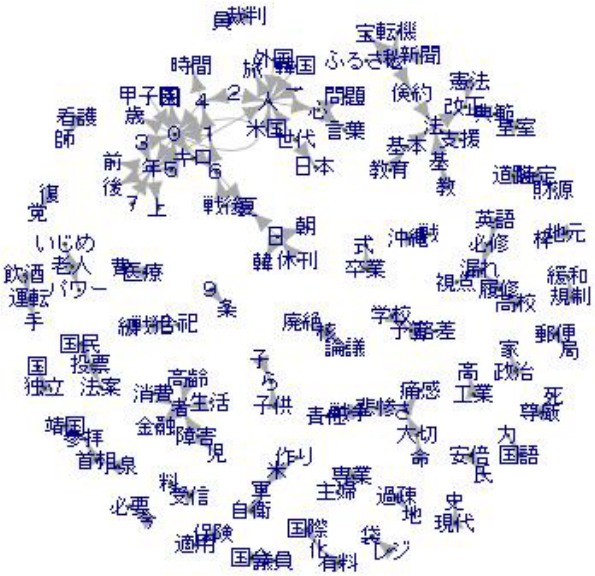

図８　「声」タイトルの抽出語ネットワーク（最小頻度３回、2006年）

さらに図８は、同じくタイトルを形態素解析した結果から、連続して共起する単語[9]を抽出し、三回以上観察されるもののみについて、単語のつながりをネットワーク表現したものである。これをみれば、ただ単に単語で内容を類推するだけでなく、言葉のつながりとして、どのような内容が語られているのか、その概要を把握することができる。例えば、左上の部分に、「老人」という単語があるが、この単語には「老人→パワー」という共起関係と、「老人→いじめ」という共起関係がそれぞれ三回以上観察されている。右下の方にある、「戦争」については、「戦争→責任」というつながりもあれば、「戦争→悲惨」というつながりもある。

こうした分析を行えば、記事中で語られているおおよそ話題が見えてくる。戦争、政治、教育といった内容がそれである。もちろん、正確な内容については、実際の記事を精査する必要がある。そういう意味で、この分析は粗っぽい概要把握である。もし、分析対象の記事が100や200程度のものであれば、それらの記事を目で読み通せば、概要把握は正確にすぐにできる。しかし、記事の総数が数千、数万の単位になってくれば、手作業での概要把握は非常な労力を要する。その場合、今回のようなツールを用いて分析をすることは、情報の効率的な整理に役立つだろう。

6 今後の課題

最後に、いくつかの課題を提示しておく。本稿では、2006年の朝日新聞の「声」に関する作業のみを紹介した。しかし、朝日新聞は1984年から、読売新聞は1987年から、毎日新聞は1991年から、それぞれ電子媒体の記事データが入手可能である。各紙それぞれ読者投稿欄が存在するので、三大全国紙の20年前後のデータを分析できることになる。これは、日本社会のある側面を切り出

9）同じタイトル中に同時に続けて出現する単語のこと。

した、貴重な社会調査データといえるだろう。ここから、例えば、「戦争」に対する日本国民の価値観や報道のされ方がどのように変化したか、というような具体的テーマについて、時間の流れの中で、新聞社の違いを比較しながら分析することができるだろう。職業属性が明示されていることを考慮すれば、職業意識の資料としても分析に値する。

また、既存の各種世論調査（国民性調査であるとか世界価値感調査であるとか）で抽出される日本人の意識や行動と、この新聞記事データから抽出されるものとの整合性を確認する作業も重要な課題である。

本稿では、そうした作業に入る前の準備段階を紹介するにとどまった。しかし、以上のような準備と整理ができれば、「データの偏り」に留意しつつも、豊かな分析が数多くできることが期待される。その具体的分析については、また別の機会に詳述したい。

文献

[1] 藤井美和・李政元・小杉考司, 2005,『福祉・心理・看護のテキストマイニング入門』, 中央法規出版.

[2] Grbich, C., 2007, *Qualitative Data Analysis*, Sage.

[3] Hertog, J., and Fan, D., 1995 “The Impact of Press Coverage on Social Beliefs: The Case of HIV Transmission,” *Communication Research*, 22 (5) :545-574.

[4] 樋口耕一, 2004a,「テキスト型データの計量的分析－２つのアプローチの峻別と統合－」,『理論と方法』19 (1): 101-115.

[5] 樋口耕一, 2004b,「計算機による新聞記事の計量的分析－『毎日新聞』にみる「サラリーマン」を題材に－」,『理論と方法』19 (2): 161-176.

[6] 石田基広, 2008,『R によるテキストマイニング入門』, 森北出版.

[7] Krippendorff, K., 1980, *Content Analysis; An Introduction to its Methodology*, Sage (=1989, 三上他訳,『メッセージ分析の技法「内容分析」への招待』勁草書房).

[8] Krippendorff, K., 2003, *Content Analysis; An Introduction to its Methodology*, 2nd edition, Sage.

[9] 見田宗介, 1965,『現代日本の精神構造』弘文堂.

[10] Nakano, Y., 2006, “War and the Netherlands: contents analysis of Japanese newspapers,” Netherlands Institute for War Documentation (eds) *Legacies of Violence: Explorations in Dutch and Japanese Research into Memory, Trauma and the Quality of Life.*

[11] 中野康人, 2009, 新聞記事抽出語集計資料集（朝日、読売、毎日）.

[12] 那須川哲哉, 2006,『テキストマイニングを使う技術／作る技術』, 東京電気大学出版局.

[13] Neuendorf, K., 2002, *The Content Analysis Guidebook,* Sage.

[14] 大隅昇・Lebart, L., 2002,「テキスト型データの多次元データ解析」, 柳井晴夫・岡太彬訓・繁桝算男・高木廣文・岩崎学編『多変量解析実例ハンドブック』, 757-783.

[15] 大谷信介, 2008,「「世論」調査の問題状況と社会調査士制度」,『社会と調査』, 1:13-22.

[16] 大塚裕子・乾孝司・奥村学, 2007,「意見分析エンジン－計算言語学と社会学の接点－」, コ

ロナ社.
[17] 盛山和夫, 2008,「社会調査にとって本当の課題はなにか」,『社会と調査』, 1:6-12.
[18] 太郎丸博, 1999,「身の上相談記事から見た戦後日本の個人主義化」, 光華女子大学文学部人間関係学科編『変わる社会・変わる生き方』ナカニシヤ出版, pp.69-93.
[19] Winship, C. and Sobel, M., 2004, "Causal Inference in Sociological Studies," Hardy (eds) *Handbook of Data Analysis* 481-503.
[20] Woodward, J.L., 1934, "Quantitative Newspaper Analysis as a Technique of Opinion Research," *Social Forces* 12:526-537.

Abstract

Japanese Newspapers as Social Survey Data: Quantitative Analysis of readers' columns

Yasuto Nakano
Kwansei Gakuin University

The purpose of this paper is to propound methods and problems of quantitative analysis of newspapers. Using newspapers as research data is a not unfamiliar way to inquire society especially in media studies. Over the past few decades a considerable number of studies have been made. Recent developments of archives and innovations of methodology make it easier to analyze Japanese newspapers in a quantitative way. Large quantity of newspaper data could be a fruitful source of social inquiry. In this paper, readers' columns of ASAHI (2006) are analyzed as a test case. The columns include not only text of content but also age and occupation of its contributor. There would be a possibility to use newspaper data as social survey data.

Keywords
newspaper, quantitative text analysis, content analysis, social survey